Review of Online LSTM Neural Network for Crypto Price Prediction

Nick Russert, Drew, Connor, Sam Braude

Abstract

Predicting the price of Bitcoin, a highly volatile cryptocurrency, remains a significant challenge in financial markets due to its complex, nonlinear, and time dependent nature. This study explores the effectiveness of advanced machine learning models for bitcoin price prediction by comparing three approaches: a Long Short-Term Memory (LSTM) network with online learning, a Convolutional Neural Network, and a simple regression model. The LSTM online learning model was designed to iteratively adapt to new data, simulating real world scenarios where models are updated incrementally based on recent information. In contrast, the CNN and regression models serve as baselines for comparison, relying on batch processing and static training. This work underscores the importance of adaptive learning in financial forecasting and highlights the advantages of integrating real time updates in predictive models. The findings contribute to the development of robust and practical tools for traders, investors and analysts navigating volatile markets like crypto.

1. Introduction

Crypto currencies, led by bitcoin, have emerged as a revolutionary force in financial markets, characterized by their decentralized nature, global accessibility, and extreme volatility. Since its inception in 2009, Bitcoin has gained widespread adoption as both a digital asset and a speculative investment. Price movements are often influenced by a complex interplay of market sentiment, regulatory developments, and macroeconomic factors. The unpredictable nature of Bitcoin and cryptocurrencies pose a significant challenge to investors, traders, and analysts who seek to make informed investment decisions.

The ability to accurately predict movements in the crypto market is not only of academic interest but also has practical applications for financial planning and risk management. Due to the nonlinear and highly dynamic nature of crypto, traditional forecasting methods can be insufficient in capturing the complex dependencies.

This study focuses on evaluating and comparing three machine learning and neural network models for Bitcoin price prediction: (1) an LSTM model with online learning, designed to simulate real world conditions by dynamically updating its knowledge with new data; (2) a CNN model. Which leverages spatial features in the time series for trend recognition; and (3) a linear regression model that serves as the baseline for comparison. By exploring the strengths and weaknesses of each approach, this research aims to provide insight into the applicability of these models for forecasting in a volatile market like crypto.

2. Related Work

The unpredictability of cryptocurrency markets has promoted extensive research into price prediction methodology. Traditional approaches for prediction such as ARIMA (autoregressive integrated moving average) and exponential smoothing, often fail to capture the non linear and high volatility of cryptocurrencies such as bitcoin. As a result, machine learning and neural network based approaches have become a dominant focus in recent times due to the ability to extract meaningful patterns from complex data.

2.1 Long Short Term Memory Model

Long short term memory (LSTM) networks have emerged as a leading approach for time series forecasting, making it a suitable network to explore in the financial domain. The ability to model long term dependencies and non linear relationships make it well suited for cryptocurrency price prediction. Wu et al. (2024) conducted an extensive review of deep learning models for crypto price forecasting, highlighting the superior performance of LSTMs and their ability to capture patterns. They further emphasize that LSTM models, in particular those with multivariate features, far outperform traditional statistical models in terms of prediction accuracy. These findings align with the goals of this study, where the LSTM online learning model is explored as a means to enhance predictive performance through continuous updates.

2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have also been explored for financial forecasting. CNNs are typically used for spatial data however they have demonstrated some ability in time forecasting due to the ability to identify patterns in structure data, such as candlestick charts and derived indicators. Wu et al. (2024) found that CNNs, when paired with additional preprocessing or feature extraction, can effectively capture broader market trends, although they often lag behind LSTMs. We use CNNs as well as linear regression as baseline methods in order to provide benchmarks for assessing the effectiveness of the more sophisticated models.

Building on these findings, this research aims to extend the scope by incorporating online learning to simulate real world conditions for a LSTM model.

3. Methodologies

3.1 Data Collection and Preprocessing

Bitcoin price data was collected from Yahoo Finance, covering a period from 2015 to 2024. Key feature engineering include:

- Moving average Convergence Divergence (MACD): captures momentum changes.
- Relative Strength Index (RSI): Measure overbought or oversold conditions.
- Logarithmic Price Transformations: ensures data normalization.

3.2 Long Short-Term Memory (LSTM)

Long short term memory (LSTM) networks are a specialized type of Recurrent Neural Network (RNN) designed to overcome the limitations of traditional RNNs in learning long term dependencies. Unlike standard RNNS, LSTMs employ a series of gating mechanisms to selectively keep or discard information over time. The architecture of an LSTM consists of:

- Input Gate: controls how much of the current input is allowed to flow into cell state
- Forget gate: determines what information from previous cell state should discard
- Output gate: regulated what part of the updated cell state is passed to the output

Key Variables and Notations

- X_t - input vector at time step t
- h_t - Hidden state vector at time step t (output of the LSTM cell)
- C_t - cell state vector at time step t
- f_t - Forget gate vector at time step t
- i_t - Input gate vector at time step t
- o_t - output gate vector at time step t
- $\overline{C}_t$ - candidate cell state at time step t
- W, U, b: weight matrices and bias terms
- σ - Sigmoid activation function
- Tanh: Hyperbolic tangent activation function.

Equations of the LSTM Cell

1. Forget Gate - Decides which information from the previous cell state C_{t-1} should be discarded:
   $f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$

2. Input Gate - Determine which new information to store in the cell state:
   $i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$

3. Candidate Cell State - Contains potential new values to add to the cell state:
   $\overline{C_t} = tanh(W_c x_t + U_c h_{t-1} + b_c)$

4. Update Cell State - The cell state C_t is updated by combining the forget gate and input gate information:
   $C_t = f_t \odot C_{t-1} + i_t \odot \overline{C_t}$
   ($\odot$ is element wise multiplication)

5. Output Gate - Determines the output for the current time step based on updated cell state:
   $o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$

6. Hidden State - Filtered version of the cell state:
   $h_t = o_t \odot tanh(C_t)$

Workflow and Summary at each time step t:

1. The model receives an input x_t and the previous hidden state h_{t-1} and a cell state C_{t-1}.
2. The forget gate f_t decides which part of C_{t-1} to retain.
3. The input gate i_t and candidate state $\overline{C_t}$ decide what new information to incorporate.
4. The cell state C_t is updated with a combination of f_t and $i_t \odot \overline{C_t}$.
5. The output gate o_t determines the final hidden state h_t, which serves as the output of the LSTM at time t.

For Crypto price prediction, the ability to capture both short term and long term trends is essential. LSTM is a clear model to explore because they can account for the volatile, high frequency nature of bitcoin and the underlying dependencies.

3.3 Online learning Implementation

Online learning refers to a learning paradigm where the model is updated as new data become available. Unlike batch learning where a model is trained on fixed data and then remains static after that, online learning allows the model to adapt dynamically and change with the new available data. This is useful in a market like crypto because of the highly volatile environment where price movements are constantly fluctuating.

4. Evaluation

4.1 Back Testing

Back testing was used in order to evaluate the effectiveness and the real world applicability of the models. This allows for the simulation of a live trading environment by using historical data in a sequential manner, allowing the models to make predictions on unseen data. This approach Incorporates:

- Buy/Sell Thresholds: models generate buy signals when the predicted price exceeds a threshold and sell signal when the predicted price decrease surpasses a negative threshold
- Stop-Loss and Take Profit: If the price movement exceeds predefined loss or profit levels, this mitigates excessive risk and loss and locks in guaranteed gains.
- Maximum Allocation: Limits the proportion of the portfolio allocated to the single trade managing risk.

Backtesting was conducted over the last 365 days of the testing dataset to simulate market conditions. The starting portfolio balance was set to $10,000, and cumulative portfolio values were tracked for each model and strategy.

4.2 Results

The cumulative returns for each model were compared to a simple buy and hold baseline, where the portfolio remains fully investing in Bitcoin through the backtesting period.

- Buy and Hold - Achieved a final balance of $20,096.94 representing a return of 100.97%.
- LSTM (static) - Achieved a final balance of $16,669.22 to $17,142.25 across threshold from 1% to 3% representing a return from 66.69% to 71.42%.
- Online LSTM - Achieved a final balance of $18,732.46 to $21,946.37 across threshold from 1% to 3% representing a return from 87.32% to 119.46%. Outperforming the buy and hold strategy at higher thresholds.

- Linear Regression - Achieved a final balance of $10,199.71 to $10,344.70 across threshold from 1% to 3% representing a return from 2.00% to 3.45%
- CNN - Achieved a final balance of $10,222.65 to $10, 817.29 across threshold from 1% to 3% representing a return from 2.23% to 8.17%

4.3 Insights

The online LSTM excelled in the volatile market due to the ability to update predictions in real time which allowed it to outperform the static models and even the buy and hold strategy. The static LSTM model still performed fairly well compared to the baseline models. This is largely due to the models ability to remember past patterns and sequences. Both linear regression and CNN fell short, highlighting the limitation of the static training. Figure 4.1 clearly shows the superiority of the online LSTM model.

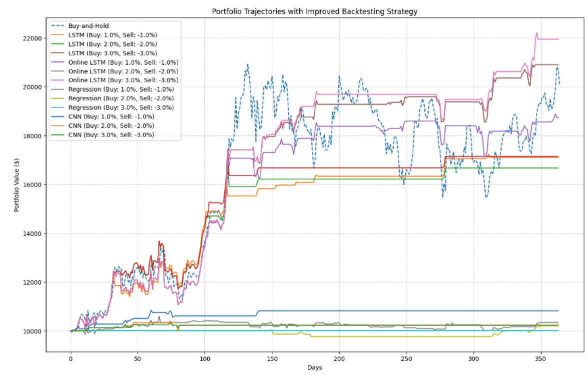

Figure 4.1

5. Conclusion

Backtesting results demonstrate that dynamic models like the Online LSTM are well suited for real world trading applications in highly volatile markets like that of cryptocurrencies. Through the incorporation of adaptive and online learning and feature engineering, these models can outperform traditional approaches and baseline strategies, providing valuable tools for traders and investors alike.

6. Efforts

Connor and Sam's main focus was code development, feature engineering, and back testing strategies for the various models. Drew and Nick focused on the presentation, researching LSTM and online learning, focusing on the paper and evaluating the results of the project.

Works Cited

Wu, J., Zhang, X., Huang, F., Zhou, H., & Chandra, R. (2024). *Review of Deep Learning Models for Crypto Price Prediction: Implementation and Evaluation*.